A Toolbox for Model-Free Analysis of fMRI Data

P. Gruber[1], C. Kohler[1], and F.J. Theis[2]

[1] Institute of Biophysics, University of Regensburg, D-93040 Regensburg
[2] Max Planck Institute for Dynamics and Self-Organization, D-37073 Göttingen

Abstract. We introduce Model-free Toolbox (MFBOX), a Matlab toolbox for analyzing multivariate data sets in an explorative fashion. Its main focus lies on the analysis of functional Nuclear Magnetic Resonance Imaging (fMRI) data sets with various model-free or data-driven techniques. In this context, it can also be used as plugin for SPM5, a popular tool in regression-based fMRI analysis. The toolbox includes BSS algorithms based on various source models including ICA, spatiotemporal ICA, autodecorrelation and NMF. They can all be easily combined with higher-level analysis methods such as reliability analysis using projective clustering of the components, sliding time window analysis or hierarchical decomposition. As an example, we use MFBOX for the analysis of an fMRI experiment and present short comparisons with the SPM results. The MFBOX is freely available for download at `http://mfbox.sf.net`.

1   Introduction

With the increasing complexity and dimensionality of large-scale biomedical data sets, classical analysis techniques yield way to more powerful methods that can take into account higher-order relationships in the data. Such often explorative methods have been popular in the field of engineering and statistics for quite some time, however they perpetrate into the application areas such as psychology, biology or medicine in a much slower fashion. This is partially due to the fact that visual and simple tools for more complex analyses are rarely available. In this contribution, we present a toolbox, MFBOX, for performing blind source separation of complex tasks with appropriate pre- and postprocessing methods. One of our main goals is to provide a simple toolbox that allows for the model-free analysis of fMRI data sets [14], although MFBOX may just as well be applied to other recordings. The graphical user interface of MFBOX enables users to easily try out various model-free algorithms, together with additional pre- and postprocessing and reliability analysis. The design of the toolbox is modular, so it can be easily extended to include your algorithm of choice. It can integrate into SPM5 [1] and can be used to perform model-free analysis of biomedical image time series such as fMRI or PET data sets. The toolbox is realized in MATLAB, and has been tested on Linux, Windows and Mac OS. The paper itself is organized as follows: In the next section, we present the core features of MFBOX,
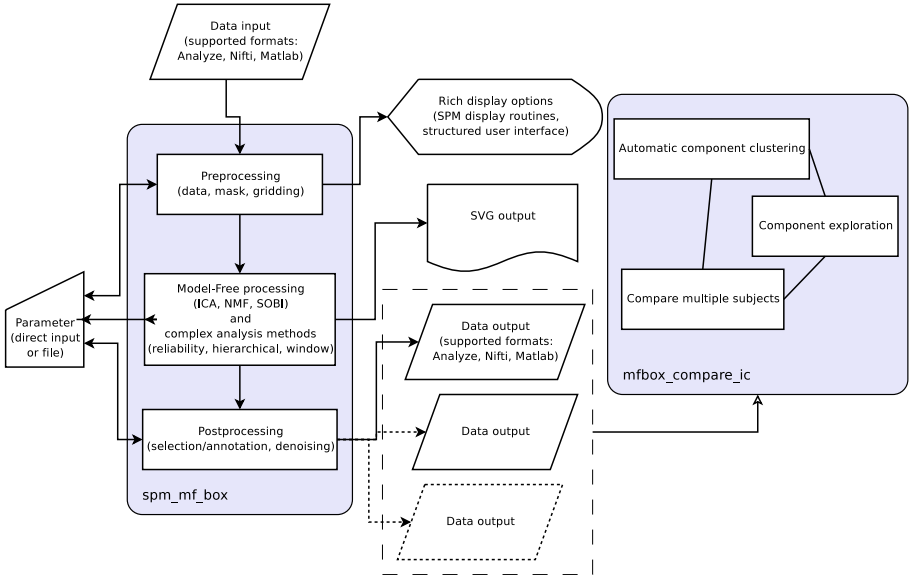
Fig. 1. The data flow of the MFBOX application. The user interface is divided into the `spm_mf_box` part and the `mfbox_compare_ic` part.

from preprocessing to Blind Source Separation (BSS) and higher-order analysis itself to postprocessing. We then illustrate the usage of MFBOX on a complex fMRI experiment, and conclude with an outlook.

2   Features of the MFBOX

The MFBOX application includes two main graphical interfaces, `spm_mf_box` which gives access to all processing possibilities and `mfbox_compare_ic` which allows for easy comparison of different results. Any of the algorithms can also be used separately from the main toolbox interface and additionally a batch run command `mfbox_runbatch` is provided to ease the analysis of multiple data sets. The workflow can be divided into modular stages, also see figure 1:

– Preprocessing
– Model-free processing and higher-level analysis
– Postprocessing

2.1   Preprocessing

The preprocessing stage can include one or more preprocessing steps in a row where the precedence can be easily controlled. The purpose of this stage is to select regions of interest to apply the model-free algorithms on or enhance the effectiveness of the main processing algorithms by denoising or transforming the original sequence data.

datavalue mask selection by bounds on voxel values
denoise high quality Local ICA (lICA) based 3d denoising [4]
infomap gridding using a Self-Organizing Maps (SOM) [12] based on the information content of the voxels
remmean different mean removal options
roiselect mask selection by loadable masks
selectslices rectangular mask selection in voxel space
varthreshold mask selection by bounds on voxel variances

Recommended standard options are to mask out the parts of the signal which are outside the brain and regions uninteresting for the Blood Oxygenation Level Dependent Contrast (BOLD) effect like blood support system. Moreover the analysis can be enhanced by only using the white matter voxels. How this mask selection is achieved depends on the available data. If structural data is available the preprocessing option roiselect can use the data from a segmentation step. Otherwise the mask selection can be accomplished by datavalue or varthreshold selection.

2.2   Model-Free Analysis and Higher-Level Analysis

Given a multivariate data set $\mathbf{x}(t)$, our goal is to find a new basis $\mathbf{W}$ such that $\mathbf{W}\mathbf{x}(t)$ fulfills some statistical properties. If we for example require the transformed random vector to be statistically independent, this is denoted as Independent Component Analysis [8]. Independent Component Analysis (ICA) can be applied to solve the BSS problem, where $\mathbf{x}(t)$ is known to be the mixture of some underlying hidden independent sources. Depending on the data set also other models are of interest, as for e.g. depending on the data set. The MFBOX currently includes three different paradigms for tackling the BSS problem and for each of these fundamental types it contains different algorithms to perform the separation.

1. ICA algorithms
   PearsonICA spatial ICA algorithm which employs a fast fixed point algorithm as extension of FastICA [8] where the nonlinearity is estimated from the data and automatically adjusted at each step [9]
   JADE spatial ICA algorithm which is based on the approximate diagonalization of the fourth cumulant tensor [3]
   stJADE spatiotemporal version of the JADE algorithm [15], for an impression on how the spatiotemporal weighting can enhance the separation of a real data set see figure 2
   TemporalICA temporal ICA [2] optionally using Projection Approximation Subspace Tracking (PAST) [17] for the temporal Principal Component Analysis (PCA) reduction
   hrfICA semi-blind spatial ICA algorithm using the Haemoglobin Response Function (HRF) function to incorporate prior knowledge about the assumed sources

2. second order algorithms

   mdSOBI Multidimensional version of the Second Order Blind Identification
   (SOBI) [16] algorithm, which is based on second-order autodecorrelation

   stSOBI Spatio-temporal version of the SOBI algorithm [15]

3. other decomposition algorithms

   hNMF a Nonnegative Matrix Factorization (NMF) decomposition algo-
      rithm using hyperplane clustering [5]

These base BSS algorithms can be combined with different types of higher-
level analysis methods. These three methods share the fact that they apply a
previously selected BSS algorithm multiple times in order to extract additional
statistics e.g. for reliability. The methods will be explained in the following.

Reliability analysis with projective k-means clustering. A common prob-
lem in explorative data analysis is how to access the reliability of the obtained
results. A proven method of reliability analysis in statistics is bootstrapping i.e.
to randomly subsample the same methods as before and compare the results
from multiple different random sub sample runs, see e.g. [6, 7]. For most BSS
algorithms this leads to a permutation problem (namely identifying correspond-
ing Independent Component (IC)) since there is an inherent permutation and
scaling, especially sign, invariance of BSS. Here we use projective k-means clus-
tering [5] for the assignment and to evaluate the quality of a component using
the resulting cluster size. It is essentially a k-means-type algorithm, which acts
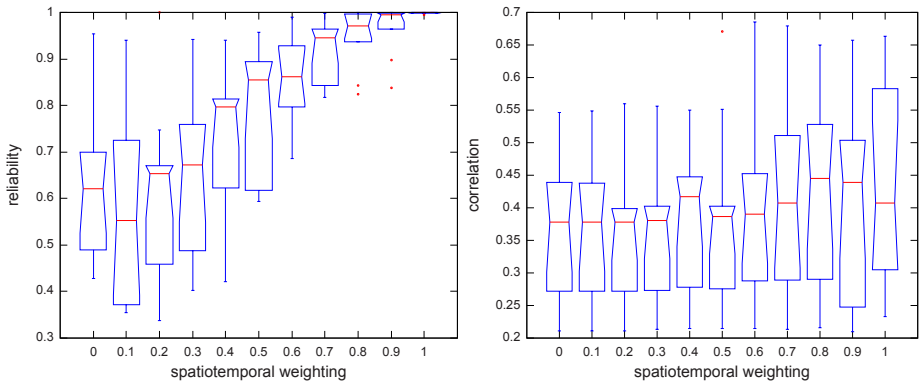
(a) Comparison of the maximal tem-
poral reliability measure with varying
weighting α

(b) Comparison of the maximal cor-
relation with the external stimulus

Fig. 2. Evaluation of some of the higher level algorithms present in the MFBOX on a
real data set using the stJADE algorithm with 10 components. In both graphs $\alpha = 1$
is the situation where only temporal ICA is performed whereas at $\alpha = 0$ only spatial
ICA is performed.

on samples in the projective space $\mathbb{RP}^n$ of all lines of $\mathbb{R}^n$. This models the scaling indeterminacy of ICA more precisely than projection onto its hypersphere since it also deals with the sign invariance. The result (figure 2) is explained by the fact that the spatial dimension of the data is strongly smoothed and reveals a some structure. Hence it does not follow the usual linear ICA model of independently drawn samples from an identically distributed random variable. The temporal dimension does not expose such additional structure and is less smoothed by the data acquisition and so the temporal ICA should be more stable.

Hierarchical analysis of component maps. Another common issue in data-driven BSS methods is how to choose the number of components and how to evaluate the large amount of possibly interesting components the process might yield without means to identify the ones which are interesting in the current problem. The hierarchical analysis tries to overcome this problem by extracting different numbers of components and extracting a hierarchical structure form the relations between the timecourses and the component maps of the extracted sources. This yields a tree structure of the components which can be used to easily navigate to the components which are of interest in an experiment. For more detailed implementation details see also [11].

Sliding time-window analysis. Usualy ICA cannot deal with non-stationary data, so most commonly approximations or windowing techniques are used. In ordinary ICA, the whole data set is transformed, and as many component maps as time steps are reconstructed. Window ICA [10] groups the observations in windows of a fixed size, and extracts components out of each of them. Correlation analysis then gives corresponding components over the whole time interval, thus yielding additional structure in a single full-scale component map.

## 2.3   Import, Export and Postprocessing

The MFBOX has rich im- and export possibilities to load and save data, masks, designs, reference brain masks, and parameters from different formats as Analyze, Nifti, plain MATLAB, and its own file format.

selectcomp semi-manual selection, labeling and grouping of extracted components
denoise lICA based 3d denoising [4] of the extracted components

3   Using the MFBOX on fMRI Recording of an Wisconsin Card Sorting Test (WCST)

In this part we will demonstrate the results from using the MFBOX on a real world data set. After a short introduction into the nature of the data we will present the result and compare it to the standard SPM result. The WCST is
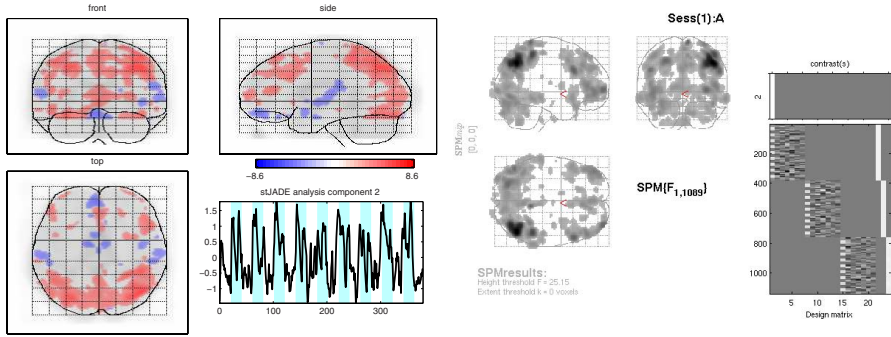
(a) The stimulus component as exported by MFBOX

(b) SPM result of the stimulus component

Fig. 3. Result of a stJADE run with α=1.0, selecting the whole brain area as mask, and after applying spatial and temporal mean removal. The component network in the prefrontal cortex and a more compact network at the parietal/occipital lobe is clearly visible.

a traditional neuropsychological test that is sensitive to frontal lobe dysfunction and allows to assess the integrity of the subjects' executive functions. The WCST-task primarily activates the dorsolateral prefrontal cortex that is responsible for executive working memory operations and cognitive control functions. The given fMRI data set originates from a modified version of the WCST [13]. Its aim is to segregate those network components participating in the above mentioned process. At first, a number of stimulus cards are presented to the participants. Afterwards the subject is asked to match additional cards to one of the stimulus cards with respect to a sorting criterion unknown to the subject. In fact, the subject has to discover it by trial and error (Sorting dimensions include: the color of symbols, their shape or the number of displayed symbols on the card). Sorting dimension changes if a previously defined number of correct answers have been given consecutively. Three different test variants of the WCST were applied

Task A. No instructions of dimensions (very close to the original WCST)

Task B. Instruction of dimensional change as sorting criterion changes.

Task C. Reminder of dimension prior to each trial; subject knows in advance the attribute that was searched for in the test.

Our results largely verify the findings of [13], as can be seen in figure 3(a). In addition to the stimulation of the prefrontal cortex, an increased activity in the parietal lobe as well as in the occipital lobe was revealed by our stJADE algorithm. The increased activity in the rear section of the brain (see figure 3(b)) was also discovered by the model-free algorithm. The complete summary of one stJADE analysis of the data set is shown in figure 4. The classification into
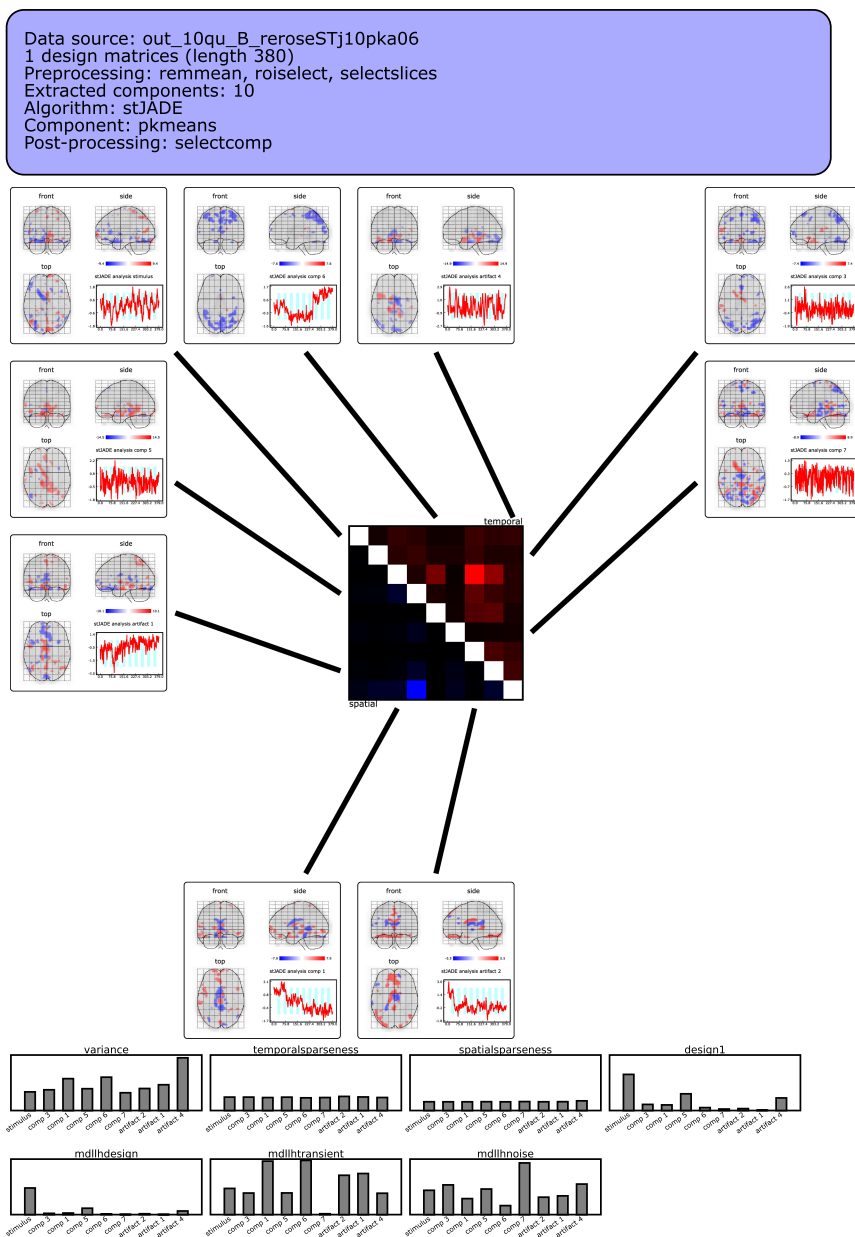
Fig. 4. A complete BSS analysis output as provided by the SVG export function of the MFBOX. The component labeled stimulus indicates the one which has the highest correlation with the stimulus. The components labeled artifact are most likely artifacts, two of them are also grouped together such that the number of displayed components is 9 although 10 components were extracted by stJADE.

stimulus component and artifacts was done after the analysis using the reliability analysis and a Minimum Description Length (MDL)-likelihood based noise estimator included in the MFBOX.

4 Conclusion

The MFBOX is an easy to use but nonetheless powerful instrument for explorative analysis of fMRI data. It includes several modern BSS algorithms and due to its highly modular structure it can easily be extended with novel as well as classical approaches.

References

[1] SPM5. edn. 5 (July 2005), `http://www.fil.ion.ulc.ac.uk/spm/spm5.html`

[2] Calhoun, V.D., Adali, T., Pearlson, G.D., Pekar, J.J.: Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. Hum.Brain Map. 13, 43–53 (2001)

[3] Cardoso, J.-F., Souloumiac, A.: Blind beamforming for non gaussian signals. IEE Proceedings-F 140(6), 362–370 (1993)

[4] Gruber, P., Stadlthanner, K., Böhm, M., Theis, F.J., Lang, E.W., Tom é, A.M., Teixeira, A.R., Puntonet, C.G., Górriz, J.M.: Denoising using local projective subspace methods. Neurocomputing 69, 1485–1501 (2006)

[5] Gruber, P., Theis, F.J.: Grassmann clustering. In: Proc. of EUSIPCO, Florence, Italy (2006)

[6] Harmeling, S., Meinecke, F., Müller, K.R.: Injecting noise for analysing the stability of ICA components. Signal Processing 84, 255–266 (2004)

[7] Himberg, J., Hyvärinen, A., Esposito, F.: Validating the independent components of neuroimaging time-series via clustering and visualization. NeuroImage 22, 1214–1222 (2004)

[8] Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis (2001)

[9] Karvanen, J., Koivunen, V.: Blind separation methods based on pearson system and its extensions. Signal Processing 82(4), 663–673 (2002)

[10] Karvanen, J., Theis, F.J.: Spatial ICA of fMRI data in time windows. In: Proc. MaxEnt 2004 AIP conference proceedings, Garching, Germany, vol. 735, pp. 312–319 (2004)

[11] Keck, I.R., Lang, E.W., Nassabay, S., Puntonet, C.G.: Clustering of signals using incomplete independent component analysis. In: Cabestany, J., Prieto, A.G., Sandoval, F. (eds.) IWANN 2005. LNCS, vol. 3512, pp. 1067–1074. Springer, Heidelberg (2005)

[12] Kohonen, T.: Self-Organizing Maps. Springer, New York, Inc. Secaucus, NJ (2001)

[13] Lie, C.-H., Specht, K., Marshall, J.C., Fink, G.R.: Using fMRI to decompose the neural processes underlying the wisconsin card sorting test. NeuroImage 30, 1038–1049 (2006)

[14] McKeown, M.J., Sejnowski, T.J.: Independent component analysis of FMRI data: Examining the assumptions. Human Brain Mapping 6, 368–372 (1998)

[15] Theis, F.J., Gruber, P., Keck, I.R., Meyer-Bäse, A., Lang, E.W.: Spatiotemporal blind source separation using double-sided approximate joint diagonalization. In: Proc. of EUSIPCO, Antalya, Turkey (2005)

[16] Theis, F.J., Meyer-Bäse, A., Lang, E.W.: Second-order blind source separation based on multi-dimensional autocovariances. In: Puntonet, C.G., Prieto, A.G. (eds.) ICA 2004. LNCS, vol. 3195, pp. 726–733. Springer, Heidelberg (2004)

[17] Yang, B.: Projection approximation subspace tracking. IEEE Trans. on Signal Processing 43(1), 95–107 (1995)